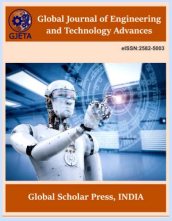
Global Journal of Engineering and Technology Advances
eISSN: 2582-5003
Cross Ref DOI: 10.30574/gjeta
Journal homepage: https://gjeta.com/



(RESEARCH ARTICLE)

Check for updates

A study and evaluation of transfer learning–based deep learning models for common skin disease detection

Nguyen Thi Bich Ngoc *

Faculty of Information Technology, Sao Do University, HaiDuong, Vietnam.

Global Journal of Engineering and Technology Advances, 2026, 26(02), 001-007

Publication history: Received on 20 December 2025; revised on 28 January 2026; accepted on 30 January 2026

Article DOI: https://doi.org/10.30574/gjeta.2026.26.2.0036

Abstract

The application of artificial intelligence in medical image–based diagnosis has attracted increasing attention, particularly in dermatology, where accurate and timely analysis of skin lesions is essential. This paper investigates and evaluates the performance of several transfer learning–based deep learning models for the automatic detection of common skin diseases. Specifically, three representative convolutional neural network architectures—ResNet50, Inception, and VGG19—are comparatively analyzed. All models are trained and evaluated on the same clinical skin image dataset, employing a unified preprocessing pipeline and data augmentation techniques to enhance generalization performance. The models are assessed using Accuracy, Precision, Recall, and F1-score as evaluation metrics. Experimental results demonstrate that VGG19 outperforms the other architectures, achieving an accuracy of 94.42%, with Precision, Recall, and F1-score all exceeding 94%, indicating stable and well-balanced classification performance. In contrast, ResNet50 and Inception achieve accuracies of 58.37% and 89.7%, respectively, reflecting their comparatively limited effectiveness under the same experimental conditions. These findings highlight the suitability of VGG19 for skin disease detection tasks in scenarios with limited data and provide practical insights for selecting appropriate transfer learning models in real-world dermatological decision-support systems.

Keywords: Skin disease detection; Deep learning; Transfer learning; VGG19

1. Introduction

The skin is the largest organ of the human body, and skin-related diseases exhibit a high prevalence worldwide. Accurate and timely diagnosis of skin lesions, particularly those with malignant potential such as melanoma, is critical for improving patient survival rates and treatment outcomes. However, dermatological diagnosis in current clinical practice faces several challenges, including the shortage of dermatology specialists in many regions, subjectivity in visual inspection, and substantial clinical similarity among different skin conditions. These limitations may result in misdiagnosis, delayed treatment, and increased burden on healthcare systems.

In recent years, deep learning has achieved remarkable progress in medical image analysis, especially in the field of dermatology. Convolutional neural networks (CNNs) have demonstrated strong capabilities in extracting complex visual features and supporting diagnostic decision-making with high accuracy, in some cases achieving performance comparable to or exceeding that of experienced dermatologists. In [1], the authors compared the performance of a convolutional neural network trained on dermoscopic images with that of 145 dermatologists in the task of melanoma classification from clinical images. The results showed that the CNN achieved performance comparable to experienced clinicians. To further improve classification performance, Li and Shen [2] proposed a CNN-based framework integrating a patch-based attention mechanism with diagnosis-guided loss weighting. This approach effectively handles high-resolution images and mitigates class imbalance, leading to improved average sensitivity across skin lesion categories.

* Corresponding author: Nguyen Thi Bich Ngoc.

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0.

Moving toward more comprehensive diagnostic systems, Liu et al. [3] developed a multimodal deep learning framework that integrates clinical images with patient data to perform differential diagnosis across 26 common skin diseases. Evaluated on 16,114 clinical cases, the system demonstrated diagnostic performance comparable to that of board-certified dermatologists. High-quality datasets also play a crucial role in advancing dermatological AI research. Finlayson et al. [4] examined the vulnerability of skin cancer classification models to adversarial attacks such as FGSM and PGD and demonstrated that adversarial training can significantly enhance model robustness against such threats. The paper [5] presents a comprehensive evaluation of deep learning–based skin lesion classification using transfer learning and uncertainty quantification on the HAM10000 dataset. Multiple pre-trained feature extractors, including CLIP variants, ResNet50, DenseNet121, VGG16, and EfficientNet-V2-Large, were benchmarked with traditional classifiers, where CLIP ViT-H/14 combined with SVM achieved the best performance. Furthermore, uncertainty quantification methods such as Monte Carlo Dropout, Ensemble, and Ensemble Monte Carlo Dropout were employed, demonstrating that ensemble-based approaches provide a favorable balance between classification accuracy and prediction reliability. In the paper [6], state-of-the-art deep neural networks were trained on large-scale public datasets, including DermNet and the ISIC Archive, while explicitly leveraging disease taxonomy to enhance classification performance. The proposed approach achieved near-human accuracy, establishing new state-of-the-art results with up to 98–99% AUC across multiple classification tasks. Li and Shen [7] also used DNN to segment lesions, extract their dermoscopic features and classify them.

This paper aims to evaluate and compare the performance of three deep learning models— ResNet50, Inception, and VGG19—for multi-class classification of common skin diseases. All models are trained and evaluated on the same dermatological image dataset following standardized preprocessing procedures. All models transfer learning is employed by initializing the models with weights pretrained on the ImageNet dataset, allowing the reuse of learned visual representations and improving training efficiency under limited medical data conditions. In addition, data augmentation techniques are applied to enhance model generalization and mitigate overfitting.

2. Some Deep Learning Models for Skin Disease Detection

2.1. ResNet50 Model

ResNet50, introduced by He et al. in 2015, is a typical deep convolutional neural network architecture based on residual learning, which effectively solves the problem of performance degradation as network depth increases [10]. The ResNet50 architecture consists of 50 weighted learning layers, built from residual blocks with skip connections, allowing information and gradients to propagate directly through multiple layers without degradation. This design allows training very deep networks with high stability and good convergence. In this study, ResNet50 was used as a transfer learning model with pre-trained weights on ImageNet to evaluate the ability to exploit deep features in the problem of classifying multiple skin disease classes.

2.2. Inception Model

The Inception architecture, first introduced by Szegedy et al. in 2014 [11]. Unlike conventional CNNs that rely on a single convolutional kernel size per layer, the Inception network employs a multi-branch structure—known as the Inception module—which performs parallel convolutions with different kernel sizes (e.g., 1×1, 3×3, and 5×5), along with max pooling. This design enables the network to capture visual features at multiple spatial scales simultaneously, making it particularly effective for complex image recognition tasks. To mitigate the computational cost associated with large convolutional kernels, Inception introduces 1×1 convolutions as dimensionality reduction layers, significantly decreasing the number of parameters while preserving representational power. Subsequent variants, such as Inception-v3, further improve training stability and performance through techniques including factorized convolutions, auxiliary classifiers, and batch normalization. In this study, the Inception model is employed as a transfer learning backbone, leveraging weights pretrained on the ImageNet dataset, use to comparative model alongside VGG19, and ResNet50.

2.3. VGG19 Model

VGG19, proposed by Simonyan and Zisserman in 2014, is a deep convolutional neural network architecture that stands out for its simple yet effective design, achieving high scores in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [12]. The model consists of 19 weighted learning layers, including 16 convolutional layers using small 3×3 filters and 3 fully connected layers. This design allows the network to learn increasingly abstract hierarchical features as the depth increases, while maintaining stability during training. VGG19 uses the ReLU activation function after each convolutional layer and applies max pooling to reduce the feature space size. Despite having a large number of parameters (approximately 143 million), VGG19 is still widely used in many computer vision and biomedical problems due to its powerful feature extraction capabilities and high generalizability, especially when combined with transfer

learning from pre-trained weights on the ImageNet dataset [12]. In this study, VGG19 was applied as a transfer learning model to evaluate the effectiveness of the deep CNN architecture in the multi-class skin disease classification problem, as well as to consider the suitability of the model for diagnostic support systems requiring high accuracy.

3. Experimental Results

3.1. Dataset Used in This Study

In this paper, the publicly available Skin Disease Dataset proposed by Biswas (2023) is utilized in its entirety, comprising a total of 1,159 labeled dermatological images spanning multiple skin disease categories [13]. The dataset is partitioned into training and testing subsets using an 80/20 split, ensuring that the class distribution is preserved across both subsets and that no overlap exists between them. Figure 1 illustrates some of the data in the paper. To enhance model generalization, mitigate overfitting, and simulate real-world variations in skin lesion appearance, online data augmentation techniques are applied during the training phase. Specifically, random transformations include image rotation within a range of ±20 degrees (rotation_range = 20), random zooming with a scaling factor of up to 20% (zoom_range = 0.2), and horizontal flipping (horizontal_flip = True). These augmentations aim to increase data diversity without altering the underlying semantic content of the images. All input images are normalized by rescaling pixel intensity values to the range [0, 1] through division by 255 (rescale = 1./255). Following preprocessing, all images are resized to a standardized resolution of 224 × 224 pixels to ensure compatibility with the input requirements of the employed convolutional neural network architectures. The data are loaded in batches of size 32 (batch_size = 32) for both training and testing phases, balancing computational efficiency and memory usage.

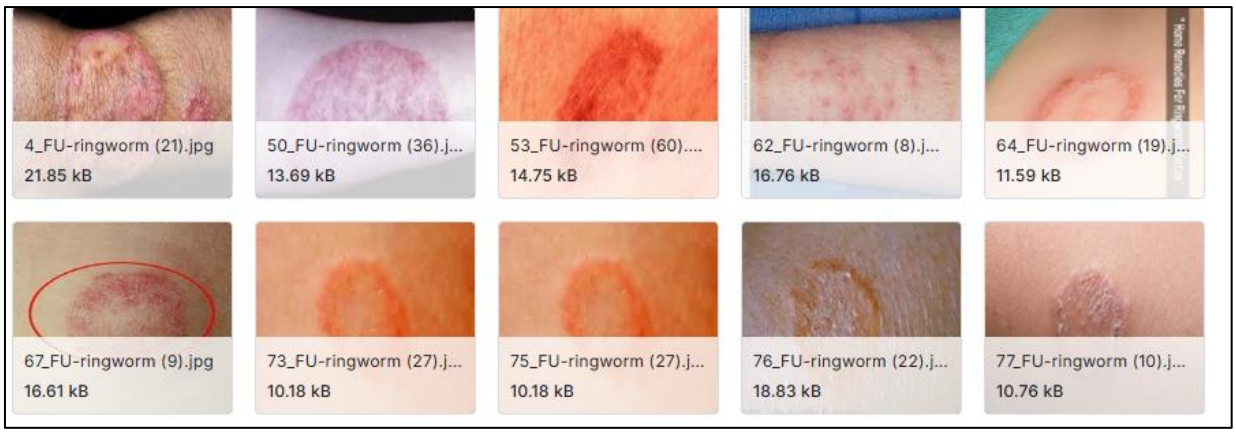


Figure 1 Some Collected Data in This Paper

3.2. Experimental Results

All experiments were conducted on the Kaggle platform utilizing an NVIDIA Tesla P100 GPU to accelerate model training and inference. The deep learning models were implemented using the TensorFlow and Keras libraries. For consistency and fair comparison, all models were trained under the same experimental settings. The Adam optimizer was employed with an initial learning rate of 1e-4, a batch size of 32, and the categorical cross-entropy loss function was used to handle the multi-class classification task.

The training behavior of the ResNet50 show in Figture 2 model indicates relatively limited performance during the initial feature extraction phase. Both training and validation accuracies remain low and increase slowly throughout the first 100 epochs, while the loss decreases only marginally. This suggests that the high-level features learned from ImageNet are not sufficiently discriminative for skin disease classification without further adaptation. A substantial improvement is observed after the fine-tuning phase begins, where deeper layers of the network are unfrozen and retrained. Training and validation accuracies increase sharply, accompanied by a significant reduction in loss values.However, noticeable fluctuations and a gap between training and validation curves indicate that ResNet50 may suffer from unstable convergence during fine-tuning. InceptionV3 show in Figture 3 exhibits faster and more stable convergence compared to ResNet50. Both training and validation accuracies rise rapidly in the early epochs and reach relatively high values, while the loss decreases steadily. This behavior highlights the effectiveness of the Inception architecture in capturing multi-scale features, which are particularly suitable for the diverse visual patterns present in skin lesion images. Nevertheless, once fine-tuning is initiated, a slight degradation in performance is observed, with a

decrease in accuracy and a corresponding increase in loss. This suggests that fine-tuning may disrupt previously learned representations if not carefully controlled. Despite this, InceptionV3 maintains strong overall performance and demonstrates good generalization capability. Among the three evaluated models, VGG19 show in Figture 4 demonstrates the most stable and consistent training behavior. Training accuracy increases steadily and reaches a very high level, while validation accuracy remains close to the training curve, indicating minimal overfitting. The loss decreases smoothly during the initial training phase and remains low throughout most of the training process. After fine-tuning, minor fluctuations are observed in the validation loss; however, the overall performance remains robust. These results indicate that VGG19 effectively adapts to the skin disease classification task when combined with transfer learning.

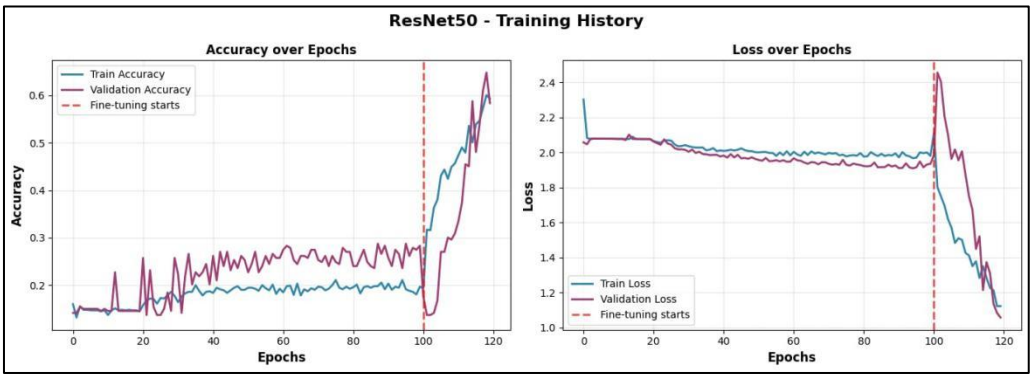


Figure 2 ResNet50 model

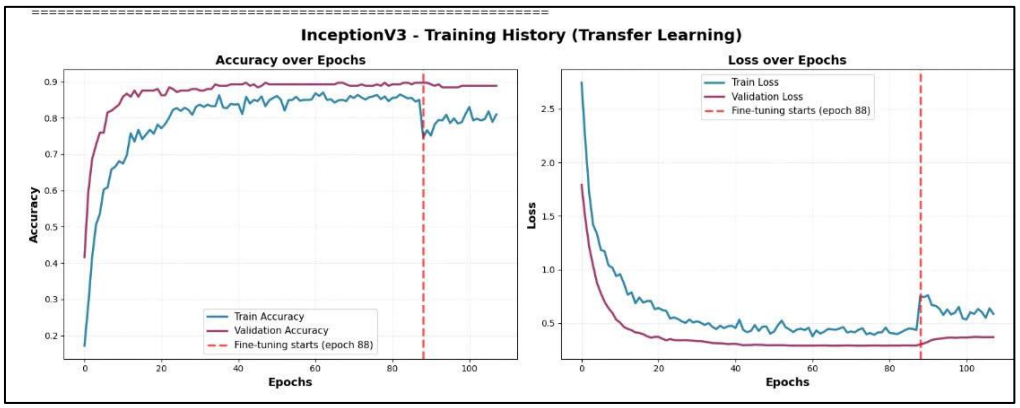


Figure 3 Inception model

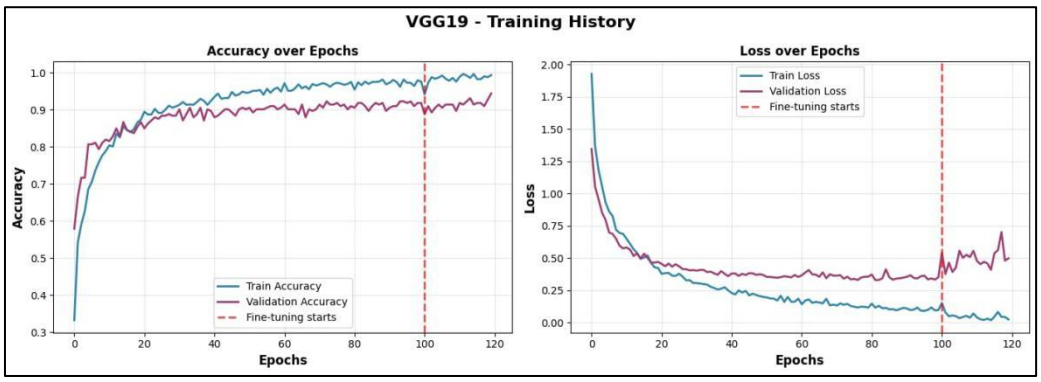


Figure 4 VGG19 model

To further investigate the class-wise performance of the evaluated models, confusion matrices were analyzed for ResNet50, InceptionV3, and VGG19. The confusion matrix of ResNet50 show in Figture 5 reveals relatively poor classification performance, which is consistent with its low overall accuracy of 58.37%. Although some classes such as

FU-athlete-foot, FU-nail-fungus, VI-chickenpox, and VI-shingles are recognized with moderate success, a substantial number of misclassifications are observed across multiple classes. Significant confusion occurs among visually similar diseases, particularly between bacterial and fungal infections. For instance, BA-cellulitis samples are frequently misclassified as fungal-related diseases, while PA-cutaneous-larva-migrans shows notable confusion with viral and fungal classes. These results indicate that ResNet50 struggles to extract sufficiently discriminative features for skin disease classification. InceptionV3 show in Figture 6 demonstrates a marked improvement over ResNet50, achieving an overall accuracy of 89.70%. Classes such as BA-impetigo, FU-athlete-foot, FU-nail-fungus, and VI-chickenpox are classified almost perfectly, indicating robust feature extraction capability. Nevertheless, limited misclassifications are still present, primarily involving FU-ringworm and PA-cutaneous-larva-migrans. These errors are likely due to overlapping visual characteristics, such as lesion shape and skin texture, which pose challenges even for experienced clinicians. Overall, InceptionV3 exhibits good generalization and significantly reduced class confusion compared to ResNet50. Among the three models, VGG19 show in Fig 7 achieves the best class-wise performance, with an overall accuracy of 94.42%. The confusion matrix displays near-perfect diagonal dominance, with minimal misclassification across all disease categories. Most classes, including BA-impetigo, FU-athlete-foot, FU-nail-fungus, VI-chickenpox, and VI-shingles, are classified with very high precision. Only minor confusion is observed in a small number of samples, particularly between PA-cutaneous-larva-migrans and fungal-related classes. However, the error rate remains very low, indicating strong discriminative power and excellent generalization ability. These results confirm that VGG19 benefits substantially from transfer learning and fine-tuning in the context of multi-class skin disease classification.

To prove the effectiveness of the proposed method, we perform statistics and compare four values after training the models including Precision, Recall, F1 Score, and Accuracy. Accuracy is the ratio between the number of correctly predicted data points and the total number of data points in the test set. Precision is the ratio of the number of correctly identified points in a class to the total number of points classified into that class. The Recall value is the ratio of the number of correctly identified points in a class to the total number of data points in that class. The quantity F1-score is the harmonic average determined based on two measures of precision and recall. The results comparing the average accuracy between the algorithms are shown in Table 1.

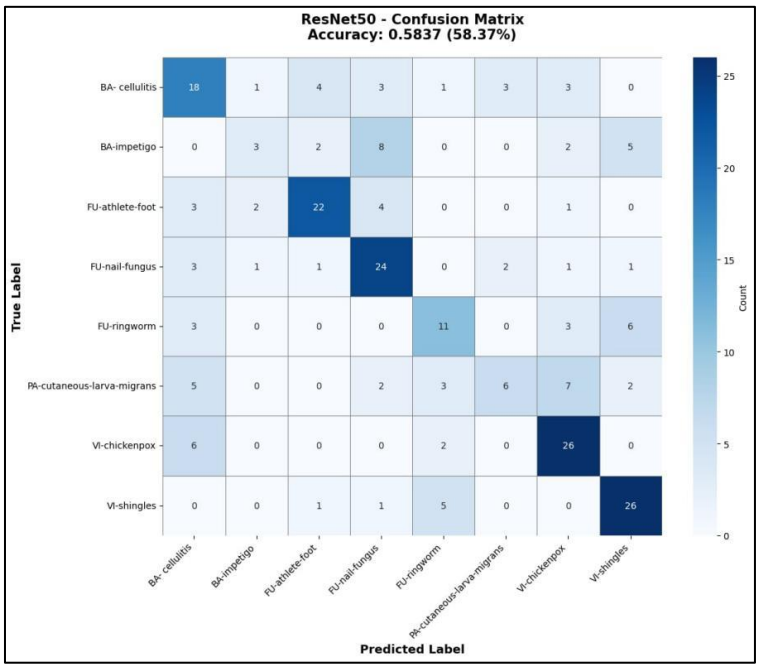


Figure 5 Confusion matrix of the ResNet50 model

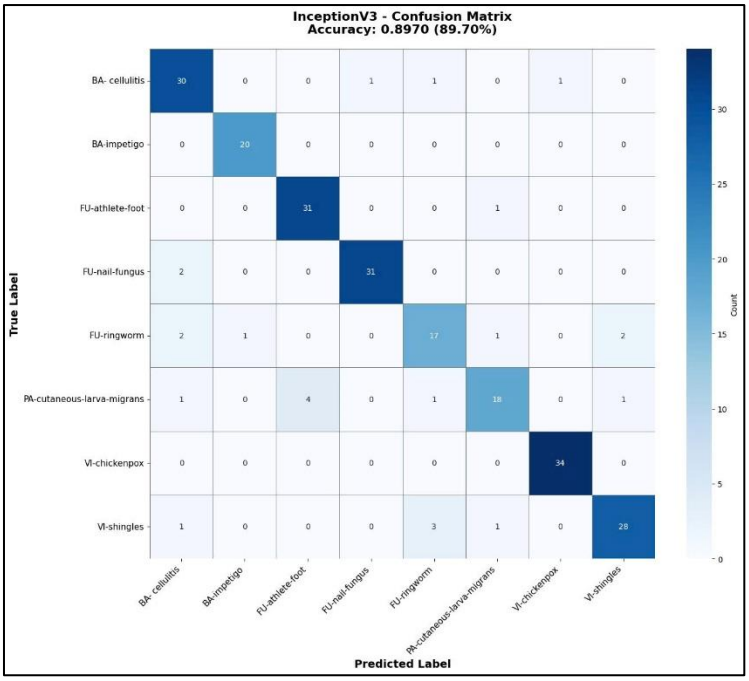


Figure 6 Confusion matrix of the Inception model

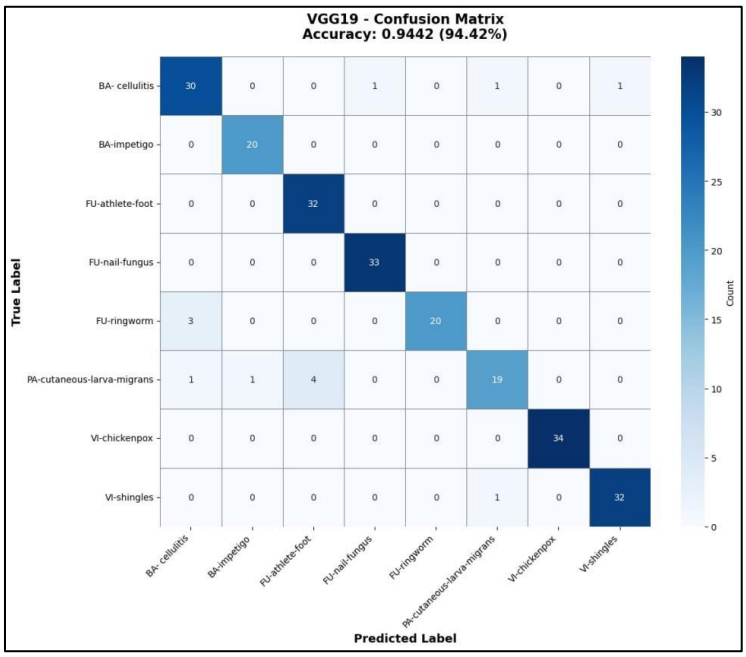


Figure 7 Confusion matrix of the VGG19 model

Table 1 Compare the accuracy between algorithms

Model	Accuracy	Precision	Recall	F1-Score
ResNet50	58.37%	57.37%	58.37%	56.25%
Inception	89.70%	89.65%	89.70%	89.54%
VGG19	94.42%	94.53%	94.42%	94.31%

4. Conclusions

This study presents a comprehensive evaluation and comparison of three representative convolutional neural network architectures—ResNet50, Inception, and VGG19—under a unified transfer learning framework for the automatic multi-class classification of eight common skin diseases based on clinical images. All models were initialized with weights pretrained on the ImageNet dataset and subsequently adapted to the target dermatological classification task through fine-tuning.Experimental results demonstrate that VGG19 achieved the best overall performance, reaching an accuracy of 94.42%, with Precision, Recall, and F1-score all exceeding 94%. This superior performance highlights the effectiveness of VGG19's deep and homogeneous architecture in leveraging pretrained visual representations and capturing subtle, discriminative features of skin lesions. In comparison, ResNet50 and Inception yielded lower classification accuracies of 58.37% and 89.7%, respectively, indicating that the benefits of transfer learning are highly dependent on the compatibility between network architecture, dataset size, and task complexity. These findings confirm the practicality of transfer learning–based deep learning approaches for dermatological image analysis , contributing toward more reliable and efficient decision-support tools in clinical practice.

References

[1] Titus J Brinker, Achim Hekler, Alexander H Enk, Joachim Klode, Axel Hauschild , Carola Berking, Bastian Schilling, Sebastian Haferkamp, Dirk Schadendorf, Stefan Fröhling, Jochen S Utikal, Christof von Kalle. "A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task". Annals of Oncology, 29(8), 1836-1842. DOI: 10.1016/j.ejca.2019.02.005

[2] Nils Gessert, Thilo Sentker, Frederic Madesta, Rüdiger Schmitz, Helge Kniep, Ivo Baltruschat, René Werner, Alexander Schlaefer. "Skin lesion classification using CNNs with patch-based attention and diagnosis-guided loss weighting". IEEE Transactions on Biomedical Engineering, 67(9), 2465-2474. DOI: 10.48550/arXiv.1905.02793.

[3] B Yuan Liu, Ayush Jain, Clara Eng, David H. Way, Kang Lee, Peggy Bui, Kimberly Kanada, Guilherme de Oliveira Marinho, Jessica Gallegos, Sara Gabriele, Vishakha Gupta, Nalini Singh, Vivek Natarajan, Rainer Hofmann-Wellenhof, Greg S. Corrado, Lily H. Peng, Dale R. Webster, Dennis Ai, Susan J. Huang, Yun Liu, R. Carter Dunn & David Coz. "A deep learning system for differential diagnosis of skin diseases". DOI: 10.1038/s41591-020-0842-3

[4] Vinay Jogani, Joy Purohit, Ishaan Shivhare, Samina Attari, Shraddha Surtkar. "Adversarial Attacks and Defences for Skin Cancer Classification". DOI: 10.48550/arXiv.2212.06822.

[5] C H. Asgharnezhad, P. Tabarisaadi, A. Khosravi, R. Alizadehsani, and U. R. Acharya, "Uncertainty-Aware Deep Learning for Automated Skin Cancer Classification: A Comprehensive Evaluation," arXiv preprint arXiv:2506.10302, 2025.

[6] M. N. Bajwa, K. Muta, M. I. Malik, S. A. Siddiqui, S. A. Braun, B. Homey, A. Dengel and S. Ahmed, "Computer-Aided Diagnosis of Skin Diseases Using Deep Neural Networks," *Applied Sciences*, vol. 10, no. 7, p. 2488, Apr. 2020, doi: 10.3390/app10072488.

[7] Li, Y.; Shen, L. Skin lesion analysis towards melanoma detection using deep learning network. *Sensors* 2018, *18*, 556.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778.

[9] Szegedy, W. Liu, Y. Jia, S. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 1–9.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in International Conference on Learning Representations (ICLR), San Diego, CA, USA, 2015.

[11] S. Biswas, "Skin Disease Dataset," Kaggle, 2023. [Online]. Available: Kaggle Dataset, doi: 10.34740/kaggle/dsv/6695743.